

<https://helda.helsinki.fi>

Kokeellinen yhteiskuntatiede

Kuorikoski, Jaakko Erkinpoika

Gaudeamus

2018

Kuorikoski , J E & Reijula , H S 2018 , Kokeellinen yhteiskuntatiede . julkaisussa T Kaidesoja
, T Kankainen & P Ylikoski (toim) , Syistä selityksiin : kausaali- ja selittäminen
yhteiskuntatieteissä . Gaudeamus , Helsinki , Sivut 279-307 .

<http://hdl.handle.net/10138/310031>

unspecified

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

KOKEELLINEN YHTEISKUNTATIEDE

Jaakko Kuorikoski & Samuli Reijula

Lisäävätkö television katselu tai tietokonepelien pelaaminen epä-sosiaalista käyttäytymistä? Miksi ihmiset oikeastaan lahjoittavat hyväntekeväisyyteen? Mikä on tehokkain tapa parantaa lasten oppimistuloksia: lasten, opettajien vai vanhempien kannustaminen? Lisääkö poliittisten vastustajien henkilökohtainen kohtaaminen todella vastakkaisten näkökantojen ymmärtämistä, parantaen näin demokraattisen päätöksenteon perusteherkkyyttä? Tiivistyykö tyhmyys joukossa? Ilmiöitä havainnoimalla voidaan löytää tukea tällaisille hypoteeseille, mutta kuten hyvin tiedetään, pelkkä havaittu yhteys kahden ilmiön välillä ei paljasta niiden välillä mahdollisesti olevan vaikutussuhteen suuntaa. Havaittu yhteys voi myös aina olla seurausta yhteisistä syistä tai toisistaan riippumattomista mutta sattumalta samankaltaisista trendeistä. Haastatteluilla ja kyselytutkimuksilla saadaan lisää tietoa siitä, kuinka ihmiset mieltävät ja kokevat kiinnostuksen kohteena olevat yhteiskunnalliset ilmiöt, mutta valitettavasti ihmiset ovat usein autuaan tietämättömiä siitä, kuinka heitä ympäröivä sosiaalinen todellisuus todella toimii, ja myös siitä, miksi he itse toimivat niin kuin toimivat.¹ Henkilökohtaisesti merkitykselliset kokemukset ja anekdootit tukevat usein täysin virheellisiä ja jopa keskenään ristiriitaisia syy-vaikutussuhteita koskevia hypoteeseja.

Kokeellinen menetelmä on tunnetusti luotettavin kausaalisen tiedon lähde luonnontieteissä. Kokeiden tekeminen ei ole kuitenkaan perinteisesti kuulunut yhteiskuntatieteiden valtavirtaan, ja eräät yhteiskuntateoreettiset ja metodologiset suuntaukset

ovat suorastaan vihamielisiä koko ajatukselle. Yhteiskuntatieteiden teoreettista hajaannusta ja näennäistä kyvyttömyyttä tuottaa kumulatiivisesti kasvavaa tieteellistä ymmärrystä onkin usein puolusteltu sillä, että luonnontieteistä poiketen yhteiskuntatieteet perustuvat vain havainnointiin ja tulkintaan. Ajatusta ihmisestä autonomisena, ainutkertaisena ja sosiaaliseen todellisuuteen holistisesti upotettuna toimijana ei ole helppo sovittaa yhteen universaaleja lainalaisuuksia testaavien laboratoriokokeiden mielikuvan kanssa. Sosiaalinen todellisuus on sotkuista, ja kaikki yritykset uuttaa siitä koeputkeen mahtuvia palasia vääristävät tutkimuskohdetta niin paljon, ettei kokeen tuloksesta voi päätellä juuri mitään hyödyllistä. Toisaalta eräs ihmistiede, psykologia, on jo pitkään ollut vakiintunut osa kokeellisten tieteiden kenttää ja kokeet ovat aina kuuluneet olennaisena osana myös sosiaalipsykologien menetelmäkokoelmaan. Ihmisen käyttäytyminen, mukaan lukien sosiaalinen käyttäytyminen, onkin usein varsin säännönmukaista ja siten kokeelliselle menetelmälle soveltuvaa, kunhan koeasetelma on loppuun asti mietitty.

Yhteiskuntatieteiden kohteena ovat tietenkin yhteiskunnalliset ilmiöt, ei yksittäinen toimija, ja kokonaisten yhteisöjen tai yhteiskuntien kokeelliseen tutkimukseen liittyy erityisiä käytännöllisiä ja eettisiä haasteita. Ajatus kokeellisen menetelmän merkityksestä myös yhteiskuntaa koskevissa pohdintoissa on kuitenkin yhtä vanha kuin itse ajatus yhteiskuntatieteistä. Yhteiskuntatieteellisen tutkimuksen kaanon pitääkin sisällään muutaman jo klassikoksi kohonneen kokeen.² Stanley Milgram kartoitti auktoriteetin rajoja osoittamalla, kuinka koehenkilöt olivat valmiita antamaan toiselle henkilölle näennäisesti hengenvaarallisia sähköshokkeja auktoriteetin käskystä.³ Philip Zimbardon tutkimusryhmän tekemässä Stanfordin yliopiston vankilakokeessa taas tutkittiin, ovatko väärinkäytökset vankilaolosuhteissa (henkinen ja fyysinen väkivalta) ensisijaisesti selitettävissä vankien ja

vanginvartijoiden persoonallisuuspiirteisiin vedoten.⁴ Yliopisto-opiskelijoiden joukosta värvätyt koehenkilöt esittivät vangin ja vanginvartijan rooleja psykologian laitoksen kellariin lavastetussa vankilassa. Zimbardon mukaan koe riistäytyi nopeasti käsistä koehenkilöiden omaksuessa roolinsa liiankin perusteellisesti ja se jouduttiin keskeyttämään ennenaikaisesti. Kokeen on katsottu osoittavan, että sosiaalinen konteksti määrittää usein käyttäytymistä persoonallisuuspiirteitä enemmän.

Nämä kuuluisat kokeet ovat kuitenkin monella tapaa poikkeuksellisia: ne perustuivat mittavaan lavastukseen ja olivat puutteellisesti kontrolloituja sekä eettisesti hyvin kyseenalaisia. Etenkin Stanfordin vankilakokeen tulokset ja niiden tulkinnat on hiljattain vakavasti kyseenalaistettu.⁵ Kokeellinen yhteiskuntatiede ei kuitenkaan edellytä näin dramaattisia koeasetelmia. Kokeiden tekeminen on arvokas apuväline yhteiskuntaa koskevien teorioiden ja mallien sosiaalista käyttäytymistä koskevien oletuksien arvioinnissa. Lisäksi erilaisten sosiaalisten instituutioiden, kuten normien, markkinamuotojen ja kannustinjärjestelmien, ominaisuuksia voidaan tarkastella kokeellisesti.

On myös tärkeää huomata, että merkittävä osa kokeellisesta tutkimuksesta tehdään tutkimuslaboratorion ulkopuolella. Niin kutsutuissa kenttäkokeissa kohteiden satunnaistaminen ja tutkimuksen kohteena oleva ”toimenpide” suoritetaan tutkimuskohteen luontaisessa esiintymisympäristössä. Jo 1920-luvulla Harold Gosnell tutki äänestyskäyttäytymistä kenttäkokeen avulla, ja etenkin viime aikoina kokeellinen menetelmä on lähestynyt esimerkiksi taloustieteen ja politiikan tutkimuksen valtavirtaa.⁶ Se on myös olennainen osa analyyttisen sosiologian metodologista ohjelmaa. Kyselytutkimuksissa käytetään yhä useammin kokeellisen tutkimuksen elementtejä (esim. satunnaistamista). Tietoverkkojen arkipäiväistyminen on tehnyt laboratoriokokeiden järjestämisen ja tulosten analysoimisen yhä helpommaksi,

ja internetin virtuaaliset yhteisöt muodostavat erityisen kätevän ja kustannustehokkaan kohteen mielenkiintoisille suuren kokoluokan kenttäkokeille.⁷ Kenttäkokeiden järjestäminen on kasvava tutkimussuunta erityisesti niillä yhteiskuntatieteiden alueilla, joilla tiedon soveltaminen politiikkasuosituksiksi on erityisen tärkeää. Hiljattain onkin esitetty vaateita, että yhteiskunnalliseen päätöksentekoon käytetyn tiedon pitäisi aina olla lääketieteellisten suositusten tavoin satunnaistetuilla kokeilla testattua.

Tarkastellaan esimerkiksi kysymystä siitä, miltä osin erot massakulttuurituotteiden suosiossa perustuvat laatueroihin ja miltä osin sattumaan. Voiko *Mona Lisan* maineen selittää vain maalauksen sisäisten ominaisuuksien perusteella ja Harry Potterin valtaisan suosion teossarjan kirjallisilla ansioilla? Kasautuvien etujen teorian mukaan merkittävä osa eroista on seurausta siitä, että alkuaan pienet, sattumanvaraiset erot tuotteiden kulutuksessa vahvistuvat sosiaalisten verkostojen ja matkimisen myötä massahittien ja unohdettujen tusinatuotteiden väliseksi valtaviksi kuiluiksi. Ehkäpä Harry Potterin suosio onkin ainakin osittain sattumaa, seurausta siitä, että riittävän suuri ja verkostoitunut joukko sattui aikoinaan poimimaan kirjan kaupan hyllyltä. Kuinka helposti ja minkä verran sattumanvarainen alkuvaiheen suosio vaikuttaa tuotteen kokonaismenestykseen? Kysymykseen vastaaminen ei onnistu pelkästään esimerkiksi myyntilukuja tarkastelemalla. Tämä johtuu siitä, että kasautuvien etujen teoria sisältää kontrafaktuaalista tilannetta koskevan väitteen: jos nyt suosittu kulttuurituote ei olisi saavuttanut tarvittavaa alkusuosiota, siitä ei koskaan olisi tullut hittiä. Voidaan toki yrittää vertailla ominaisuuksiltaan samankaltaisten kulttuurituotteiden suosiota historiallisesti, mutta ongelmana on määrittää toteutuneesta suosiosta riippumattomat ja objektiiviset kriteerit sille, milloin kulttuurituotteet ovat sisäisiltä ominaisuuksiltaan samanlaiset.

Matthew Salganik, Peter Dodds ja Duncan Watts päättivät tutkia asiaa kokeellisesti. He loivat verkkoon keinotekoiset ”kulttuuri-markkinat”: verkkosivuston, jolla kävijät kuuntelivat ja arvostelivat tuntemattomien yhtyeiden kappaleita. Verkkosivulla vierailevat (yht. n. 14 000 osallistujaa) arvottiin heidän tietämättään kahteen ryhmään: toiset saivat informaatiota kappaleiden suosioista ja toiset eivät. Ne, jotka päätyivät sosiaalista informaatiota sisältävään ryhmään, jaettiin tämän jälkeen vielä satunnaisesti kahdeksaan itsenäiseen sosiaaliseen maailmaan. Kunkin maailman osallistujat näkivät ainoastaan tuon maailman sisäiset kappaleiden toistomäärät. Näin tutkijat pystyivät luomaan kontrolloituja vaihtoehtoisia historioita, joiden avulla voitiin tutkia sitä, vahvistaako kasautuvien erojen mekanismi alun perin pieniä ja usein satunnaisia eroja kappaleiden suosion välillä. Kokeen tulokset tukevat kasautuvien etujen hypoteesia: erot kappaleiden välisessä suosiossa olivat huomattavasti suurempia sosiaalista informaatiota sisältävissä ryhmissä, ja vaikka kappaleen ”laadulla” vaikutti olevan jonkin verran vaikutusta sen menestykseen, etenkin sosiaalisen informaation ryhmissä oli hyvin vaikea ennustaa kappaleen saavuttamaa suosiota: kappale, josta tuli hitti yhdessä sosiaalisessa maailmassa, saattoi menestyä toisessa ryhmässä heikosti. Sosiaalisella ympäristöllä ja vertaisryhmän käyttäytymisellä vaikuttaa siis olevan suuri vaikutus kulttuurituotteiden laadun arviointiin.⁸

Mikä oikeastaan tekee kokeellisesta menetelmästä niin erinomaisen? Sitä pidetään usein jopa koko uuden ajan tieteellisen vallankumouksen tärkeimpänä osana. Tällöin ajatellaan, että tiede perustuu olennaisesti hypoteettis-deduktiiviseen järjelyyn ja että kokeellinen menetelmä on hypoteesin testaamisen luotettavin muoto. Kokeellista menetelmää pidetään myös usein parhaana tai jopa ainoana tapana saavuttaa luotettavaa tietoa kausaalisuhteista. Kun tutkija voi tarkasti kontrolloiduissa laboratorio-olosuhteissa suoraan manipuloida kiinnostuksen kohteena olevaa muuttujaa,

ovat kaikki tutkimuksen kohteessa havaittavat muutokset mitä todennäköisimmin kausaalisia seurauksia nimenomaan tuosta manipulaatiosta eivätkä kausaalipäättelyä sekoittavista, havaitsemattomista yhteisistä syistä.

Tässä luvussa tarkastelemme hypoteesien testaamista ja kokeellista kausaalista järkeilyä tieteenfilosofisesta näkökulmasta. Arvioimme kokeellisen menetelmän mahdollisuuksia ja rajoituksia yhteiskuntatieteellisen tutkimuksen kontekstissa, jossa luonnontieteille ominaisia yleispäteviä teorioita harvoin on saatavilla ja jossa suoraviivaisiin kausaaliväitteisiin suhtaudutaan usein epäillen. Tämä luku ei siis ole menetelmäopas, joka kädestä pitäen neuvoisi, kuinka yhteiskuntatieteellisiä kokeita tulisi rakentaa, vaan katsaus niihin perustaviin metodologisiin kysymyksiin ja periaatteisiin, joihin varsinaiset menetelmät nojaavat.

Seuraavaksi tarkastelemme niitä perustavia ominaisuuksia, jotka antavat kokeille niiden tiedollisen erityisaseman tutkimushypoteeseja testattaessa sekä kausaalisessa päättelyssä. Tuonempana esittelemme erilaisia koetyyppejä sekä erittelemme laboratoriokokeiden ja kenttäkokeiden heikkouksia ja vahvuuksia. Lopuksi tarkastelemme yhteiskuntatieteellisiin kokeisiin liittyviä erityiskysymyksiä.

KOKEIDEN LOGIIKKA: HAVAINTOAINEISTO, ILMIÖT JA ANKARA TESTAAMINEN

Tutkimuksen katsotaan olevan kokeellista, kun tutkija kontrolloi tutkimuksen kohteena olevan ominaisuuden kohdentumista tutkittavana olevassa populaatiossa.⁹ Tyypillisesti tutkitut havaintoyksiköt jaetaan satunnaisesti ainakin kahteen ryhmään, ja toisin kuin verrokkiryhmään koeryhmään kohdistetaan interventio, toimenpide, jonka vaikutusta selitettävään muuttujaan halutaan tutkia. Esimerkiksi tutkittaessa uuden lääkkeen vaiku-

tuksia koeryhmä saa lääkettä ja verrokkiryhmä ainoastaan lueläkettä. Tutkija ei siis vain passiivisesti havainnoi, mitä tapahtuu, vaan puuttuu aktiivisesti tutkimiensä tapahtumien kulkuun paljastaakseen kohteestaan uusia piirteitä tai parantaakseen päätelmiensä luotettavuutta. Tässä aluvuossa tarkastelemme kokeiden tekemisen logiikkaa yleisellä tasolla ja tarjoamme tieteenfilosofisen jäsennyksen siitä, mikä tekee kokeellisesti tuotetusta tiedosta erityislaatuista.

Tieteenfilosofiassa kokeiden tekemisen tiedollista arvoa on perinteisesti jäsennetty hypoteettis-deduktivistisen tieteellisen päättelyn mallin (HD-malli) valossa. HD-mallin mukaan tiede edistyy, kun teorioita testataan vertaamalla niistä johdettuja ennusteita havaintoihin. Koska kaikkia teorian alaan kuuluvia mahdollisia asiantiloja ei valmiiksi esiinny luonnossa, tulee tieteilijän pakottaa luonto paljastamaan, käyttäytykö se myös tällaisessa tilanteessa teorian ennustamalla tavalla. On siis aika suorittaa koe. Koska kokeessa tuotetusta asiantilasta ei ole ennestään havaintoja, on mahdotonta, että teoria olisi räätälöity varta vasten niiden mukaiseksi, ja näin teoria tulee todella testatuksi. Usein ajatellaankin, että aidon ennusteen toteutuminen antaa vahvempaa tukea teorialle kuin jo valmiiksi tunnettu empirinen ilmiö, joka on yhteensopiva teorian kanssa.

Peruslogiikasta seuraa, että yhteensopivuus olemassa olevien havaintojen kanssa ei oikeastaan voi koskaan todistaa teoriaa todeksi, mutta teoriasta johdettujen väitteiden kanssa ristiriidassa olevat havainnot kumoavat teorian – ainakin periaatteessa. Mutta mikäli kokeen lopputulos ei vastaa teorian ennustetta, voi vika olla teorian sijaan myös koeasetelmassa. Teorioista voidaan johtaa havaintoja vastaavia ennusteita ainoastaan lukuisten apuhypoteesien avulla. Näin ollen ristiriita teoreettisen ennusteen ja havaintojen välillä itsessään oikeuttaa päättelemään vain, että jossain on vikaa, mutta ei sitä, mihin syyttävän sormen tulisi osoittaa. Tästä

syystä hypoteesien kokeellinen kumoaminen ei ole yksioikoinen toimenpide. Esimerkki falsifikationistisen päättelyn vaikeudesta ovat CERNissä vuonna 2011 tehdyt kokeet, joissa havainnoidut neutriinot vaikuttivat liikkuvan valon nopeutta suuremmalla nopeudella, näin kyseenalaistaen suhteellisuusteorian antaman ylärajan luonnossa havaittaville nopeuksille. Pian kävi kuitenkin ilmi, että tulokset eivät olleet toistettavissa ja ne todennäköisesti aiheutti huono liitos koeasetelman valokuitukaapelissa.¹⁰ Yllättävät havainnot eivät johtuneet suhteellisuusteorian virheellisyydestä vaan virheellisestä koeasetelman toimintaa koskevasta apuoletuksesta. Huomiota, jonka mukaan hypoteesit kohtaavat havainnot aina joukkona ja teorian yksiselitteinen kumoaminen ei näin ole mahdollista, kutsutaan yleisesti Duhem–Quine-teesiksi.

Tämä oppikirjamainen esitys kokeellisesta hypoteesien testauksesta on kuitenkin monin tavoin harhaanjohtava. Ensinnäkin teorioista on käytännössä mahdotonta johtaa *yksittäisiä havaintoja* koskevia täsmällisiä ennusteita edes apuhypoteesien avulla. Luonnontieteelliset teoriat tekevät ennusteita hiukkasten käyttäytymisestä magneettikentässä, populaatioiden muutoksista ympäristön muuttuessa ja isotooppien suhteellisista määristä kivilajeissa. On tärkeää huomata, että mitkään näistä *ilmiöistä* eivät ole paljain silmin yksiselitteisesti havaittavia. Tieteellisten kokeiden tuloksia ei yleisesti havaita suoraan, vaan kokeita tekeillä tuotetaan aktiivisesti *havaintoaineistoa* (dataa), jonka toivotaan kertovan jotain kokeen kohteena olevasta ilmiöstä. Valokuvat kuplakammiossa esiintyvistä kuplavanoista ovat havaintoaineistoa hiukkasten liikkeistä, ja esimerkiksi rengastetuista linnuista tehdyistä havainnoista koostetut tilastot ovat havaintoaineistoa lintupopulaatioista. Mikään tieteellinen teoria ei ennusta kaikkia yksittäisten kuplakammiovalokuvien tai rengastustilastojen ominaisuuksia (kuten yksittäisen alfahiukkasen lentorataa tai maaliskuun toisena päivänä havaittujen varpushaukkojen lukumäärää),

koska havaintoaineistossa on aina satunnaisista häiriötekijöistä aiheutunutta hälyä, joka ei varsinaisesti ole relevanttia testattavalle teorialle.¹¹ Sama erottelu ilmiöiden ja havaintoaineistojen välillä pätee myös yhteiskuntatieteissä: selittämisen ja teoretisoinnin kohteina eivät ole pörsseistä ladatut arvopaperihintatiedostot, säännönmukaisuudet translitteroiduissa keskustelunauhoituksissa tai vanhat sanomalehdet vaan ne rahoitusmarkkinoiden, sosiaalisen vuorovaikutuksen tai poliittisen legitimaation piirteet, joista näiden havaintoaineistojen avulla voidaan tehdä päätelmiä.

Ilmiöiden ja havaintoaineiston käsitteellinen erottelu on tärkeää, koska päättelyaskeleet teorian ja ilmiön ja toisaalta ilmiön ja havaintoaineiston välillä ovat ratkaisevasti erilaisia. Teoriasta voidaan johtaa ilmiön ominaisuuksia, joiden havaitseminen tutkimuksessa antaa tukea teorialle. Ilmiöiden ominaisuuksien havainnointi vaatii kuitenkin työtä. Havaintoaineiston tuottaminen ilmiöstä on kausaalinen prosessi, ja päättely havaintoaineistossa mahdollisesti esiintyvistä systemaattisista piirteistä ilmiön todellisiin ominaisuuksiin koskee tämän prosessin luotettavuutta. Tässä päättelyssä pyritään signaalin ja hälyn luotettavaan erotteluun ja virhekontrolliin. Tämäkin päättely on siis kausaalista – seikka, jonka muodolliseen logiikkaan nojannut tieteenfilosofia pitkään sivuutti. Molemmat päättelyaskeleet, ilmiöstä teoriaan ja havaintoaineistosta ilmiöön, edellyttävät lukuisia apuoletuksia. Duhem–Quine-teesi kuitenkin jättää huomiotta sen, että näiden kahden erilaisen päättelyaskeleen apuoletukset ovat usein keskenään erilaisia. Erityisen olennaista on, että oletukset, jotka mahdollistavat päättelyn havaintoaineiston piirteistä ilmiön ominaisuuksiin (esimerkiksi valokuitukaapeleiden kytkentöjen virheettömyys, se, että osakehintatiedostoihin ei ole eksynyt koodauksesta aiheutuneita systemaattisia virheitä tai se, että tutkittavan keskustelun nauhoitus ei sulje pois vuorovaikutukselle olennaisia ääniteitä) ovat yleensä riippumattomia ilmiötä selittävästä ja siten testattavana

olevasta teoriasta. Tarkastelemme seuraavaksi yksinkertaista HD-mallia tarkempaa tapaa hahmottaa kokeellisen päättelyn logiikka.

Havaintoaineiston, ilmiöiden ja taustateorioiden erottelu sopii hyvin yhteen HD-mallille vaihtoehtoisen teorioiden testaamista koskevan näkemyksen kanssa. Muun muassa Deborah Mayon muotoilema *ankaran testaamisen* ajatus pyrkii esittämään havaintoaineiston ja teorian suhteen edellä esitettyä klassista näkemystä hienovaraisemmin.¹² Myös Mayo korostaa, että tieteessä evidenssin ja tutkittavan hypoteesin suhdetta ei tulisi ajatella yleisen tason loogisena relaationa. Kun tutkija arvioi, tukeeko havaintoaineisto hypoteesia, on hän tyypillisesti kiinnostunut siitä, millaisilla menetelmillä havaintoaineisto on tuotettu ja kuinka luotettavasti se näin ollen tukee hypoteesia. Ankaraa testaamista korostava lähestymistapa siirtääkin huomiota laajamittaisten teorioiden tarkastelusta kohti yksittäisten teoreettisten väitteiden ja niiden tueksi hankitun havaintoaineiston välisen suhteen yksityiskohtaisempaa tarkastelua.

Ankaran testaamisen ajatus antaa nyrkkisäännön, jota teoreettista hypoteesia testattaessa tulisi noudattaa. Ilmiötä koskevaa teoreettista väitettä testatessa tulisi hankkia sellaista *havaintoaineistoa, jolla olisi mahdollisimman suuri todennäköisyys osoittaa testattava teoreettinen väite virheelliseksi, mikäli se sitä on*. Monimutkaiselta kuulostava ajatus osoittautuu lopulta ymmärrettäväksi, kun sen kääntää toisin päin: testin tuottama havaintoaineisto ei ole vahvaa todistusaineistoa hypoteesin puolesta, mikäli testi ei ole tehokas tapa havaita hypoteesissa mahdollisesti piileviä virheitä. Kyse on siis pohjimmiltaan Karl Popperin perusopetuksesta: tieteessä tärkeää on nimenomaan hypoteesin mahdollinen kumoaminen, ei mikä tahansa havaintoaineisto, joka on vain *yhteensopivaa* hypoteesin kanssa. Esimerkiksi edellä mainitun Stanfordin vankilakokeen päämääränä oli hankkia sellaista havaintoaineistoa, jolla voitaisiin luotettavasti testata väitettä, jonka mukaan

vankiloiden huonot olot aiheutuvat vankien ja vartijoiden ongelmallisista persoonallisuuspiirteistä. Pelkästään havainnointiin perustuva aineisto, jossa vertaillaan vankien ja ”normaalien” henkilöiden persoonallisuuspiirteitä ja asenteita, ei ole ankara testi tälle hypoteesille: vaikka vankien ja muun populaation välisissä psykologisissa profileissa havaittaisiin eroja, ovat tällaiset havainnot todennäköisiä myös tilanteissa, joissa hypoteesi ei ole tosi. Erot ryhmien psykologisten profilien välillä voivat nimittäin joutua vaihtoehtoisesta prosessista – nimenomaan altistumisesta vankilaympäristölle.

Ankaran testaamisen ajatus tarjoaa hyvän tavan selvittää kokeilla tuotetun havaintoaineiston erityislaatuisuutta. Koeasetelmia suunniteltaessa pyritään poissulkemisella tai satunnaistamisella kontrolloimaan häiritseviä tekijöitä siten, että voidaan olla varmoja siitä, että kokeessa havaittu yhteys kahden ilmiön välillä ei aiheudu taustatekijöiden (esimerkiksi yhteisten syiden) vaikutuksesta. Tällaisella testillä saatu havaintoaineisto täyttää ankan testaamisen kriteerin: mikäli hypoteesin esittämä kuva tutkittavien tekijöiden välisestä suhteesta olisi virheellinen, olisi hypoteesin kanssa sopusoinnussa olevan havaintoaineiston saaminen ollut hyvin epätodennäköistä. Kuten edellä mainitusta esimerkistä käy ilmi, ei-kokeellisen havaintoaineiston tapauksessa taas virheelisen päättelyn todennäköisyys on huomattavasti suurempi. Näin ankan testaamisen idea siis antaa perustelun sille, miksi uusi, vielä havaitsematon kokeellinen havaintoaineisto on usein tiedollisesti arvokkaampaa kuin pelkkään passiiviseen havainnointiin perustuva aineisto. Samaa tapaan testin ankaruuden arviointi selittää, miksi uutta evidenssiä tyypillisesti pidetään jo olemassa olevaa havaintoaineistoa arvokkaampana: hypoteesi voidaan aina periaatteessa *post hoc* räätälöidä yhteensopivaksi olemassa olevan havaintoaineiston kanssa, kun taas uuden havaintoaineiston ja olemassa olevan hypoteesin sopusointu olisi suuri sattuma, mikäli

hypoteesi ei esittäisi oikeansuuntaisia väitteitä tarkasteltavasta ilmiöstä. Sinällään olennaista ei kuitenkaan ole hypoteesin muotoilemisen ja havaintoaineiston hankkimisen ajallinen järjestys vaan se, että mikäli kyse olisi vain sattumasta, olisi hyvin epätodennäköistä, että uusissa oloissa kerätty havaintoaineisto käyttäytyisi hypoteesin mukaisesti.

Tästä seuraa, että myös jo olemassa oleva havaintoaineisto voi täyttää ankaran testaamisen kriteerit, jos sen on synnyttänyt kokeenkaltainen kausaalinen prosessi. Emme kuitenkaan tässä käsittele tähän mahdollisuuteen nojaavia kvasikokeita ja luonnollisia kokeita.

On muistettava, että arviot testin ankaruudesta tehdään aina teoreettisten oletusten valossa. Esimerkiksi väite, jonka mukaan havaitun säännönmukaisuuden aiheuttaa nimenomaan hypoteesissa esitetty kausaalisuhde, perustuu aina arvioon siitä, että muut mahdolliset kilpailevat selitykset eivät sovellu tarkasteltavaan tapaukseen. Useissa tapauksissa vaihtoehtoisten selityksien joukko on avoin, ja siksi tällainen arvio on väistämättä epävarma. Tieteellisen menetelmän keskeinen osa on juuri sellaisen havaintoaineiston tuottaminen, joka mahdollisimman tehokkaasti sulkee pois keskenään vaihtoehtoisia hypoteeseja. Taustamuuttujien ja koemuuttujan systemaattinen manipulointi pyrkii paikallistamaan havaittujen muutosten lähteen yhteen tiettyyn tekijään. Näin testattavan hypoteesin ollessa virheellinen todennäköisyys saada ennustettu tulos on mahdollisimman pieni. Hyvin suunnitellun kokeen tunnuspiirteitä varsinaisen testin ankaruuden ohella on myös se, kuinka helpoksi se tekee ankaruuden arvioimisen. Esimerkiksi edellä mainittua Stanfordin vankilakoetta ei enää pidetä kovin vakuuttavana, koska on käynyt ilmi, kuinka paljon Zimbardo itse vaikutti suorasti ja epäsuorasti koehenkilöiden käyttäytymiseen. Koehenkilöiden käyttäytyminen ei siis todennäköisesti seurannut spontaanisti annetuista institutionaalisista rooleista, vaan

he pyrkivät käyttäytymään kokeentekijän odotusten mukaisesti. Lisäksi koehenkilöiden vähäinen lukumäärä ja heidän käyttäytymisensä yhdenmukaisuuden räikeä liioittelu lisäävät epävarmuutta siitä, testasiko koe oikeastaan mitään hypoteesia. Juuri vankilakokeen monimutkaisuuden vuoksi on hyvin vaikea arvioida, mitä (jos mitään) koe oikeastaan osoittaa ja kuinka luotettavaa päättely kokeesta käyttäytymiseen todellisissa vankiloissa on.

Kuuluisuudestaan ja intuitiivisesta vetoavuudestaan huolimatta vankilakoetta ei siis tule pitää esimerkkinä hyvin suunnitellusta yhteiskuntatieteellisestä kokeesta. Tarkastellaan vertailun vuoksi viimeaikaista koetta, jossa pureudutaan syvemmälle kasautuvien etujen teoriaan. Edellä mainittu kulttuurimarkkinakoe osoitti kasautuvien etujen merkittävän vaikutuksen, mutta jätti täysin avoimeksi sen, mitkä mekanismit toteuttavat tämän efektin. Vakavasti otettavia mahdollisuuksia on ainakin kaksi: ihmiset käyttävät statusta (suosiota) evidenssinä tuotteen laadusta, tai ihmiset ovat ensisijaisesti kiinnostuneet statuksesta itsestään (joko koordinaatio- tai signaalointitarkoituksiin). Näiden mekanismien vaikutuksia on vaikeaa, ellei mahdotonta, erotella nojautumalla ainoastaan havaintoihin. Shelley Correll kumppaneineen testasi, riittääkö pelkästään jälkimmäinen mekanismi kasautuvien etujen syntyyn jopa sellaisessa tilanteessa, jossa havaitun statuksen tiedetään olevan ristiriidassa ihmisten aitojen preferenssien kanssa. Koeasetelma oli hyvin yksinkertainen: koehenkilöt asetettiin tilanteeseen, jossa heidän täytyi valita, viedäkö lahjaksi pienempi määrä kalliimpaa ”korkean statuksen” suklaata vai suurempi määrä halvempaa ”matalan statuksen” suklaata. Lahjan julkinen status oli siis yleisesti tiedossa, mutta koehenkilöt arvottiin ryhmiin, joissa toisille kerrottiin, että lahjan saaja todella pitää kalliimmasta suklaasta, ja toisille, että saaja itse asiassa pitää enemmän halvemmasta vaihtoehdosta. Näin sekä lahjan julkinen status että lahjan saajan todelliset mieltymykset olivat koehenkilöiden

tiedossa, joten koeasetelma sulki pois ensimmäisen kasautuvien hyötyjen mekanismin (eli sen, että suosiota olisi käytetty evidenssinä sille, kumpi lahja oli todella ”parempi”). Tästä huolimatta merkittävä osa ihmisistä valitsi korkean statuksen lahjan, vaikka he tiesivät, että lahjan saaja pitää enemmän matalan statuksen lahjasta. Kysyttäessä koehenkilöt kertoivat valinneensa korkean statuksen lahjan signaloidakseen, hieman ristiriitaisesti, lahjan saajaa kohtaan tuntemaansa arvostusta.¹³

Edellä kuvattu koe on hyvin riisuttu, ja siitä puuttuu vankilakokeen puhutteleva dramatiikka. Juuri kokeen yksinkertaisuuden vuoksi sen tulosten tulkinta ja luotettavuuden arviointi on paljon helpompaa. Koeasetelma on mahdollisimman yksinkertainen, testattava mekanismi selkeästi määritelty (testi- ja kontrolliryhmiä erottaa systemaattisesti ainoastaan yksi seikka) ja havaittujen tulosten vaihtoehtoisia selityksiä on suljettu tehokkaasti pois (mm. siten, että kaikki relevantti informaatio oli koehenkilöiden saatavilla, pyytämällä koehenkilöitä selittämään käyttäytymistään ja satunnaistamisella). Vaikka valinta erilaisten suklaalahjojen välillä ei ensi silmäyksellä vaikuta yhtä yhteiskunnallisesti ja yhteiskuntatieteellisesti tärkeältä tutkimuskohteelta kuin sosiaalisten roolien vaikutus käyttäytymiseen vankiloissa, on yksittäisestä koeasetelmasta yleistettävissä oleva sosiaalinen mekanismi eittämättä merkittävässä roolissa useiden tärkeiden statushierarkioiden ylläpitämisessä.

KOKEELLINEN MENETELMÄ JA KAUSAALINEN PÄÄTTELY

Edellä on käsitelty tieteellisiä kokeita yleisesti hypoteesien testaamisen menetelmänä. Kokeellinen menetelmä on myös luotettavin tapa tuottaa tietoa nimenomaan kausaalisuhteista. Jaakko Kuorikosken luvussa esittämän ajatuksen mukaisesti kausaalisuhteet tulee mieltää manipulaation mahdollistaviksi riippuvuus-

suhteiksi: manipuloimalla syytä voimme vaikuttaa seuraukseen, kun taas seurauksien manipuloinnilla ei voi vaikuttaa syihin. Kuten Kuorikoski luvussa esittää, voidaan kausaliiteille relevanttia manipulaation käsitettä edelleen selventää intervention käsitteellä: kausaalisuhteet ovat ideaalisesti tarkkojen ja ulkoisesti määräytyneiden manipulaatioiden suhteen invariantteja riippuvuussuhteita. Kun kausaalisuhteen käsite ymmärretään tällä tavoin, kausaalinen päättely koskee siis jo lähes määritelmällisesti ideaalisten kokeiden mahdollisia seurauksia.

Peruste yhteiskuntaa koskevan kausaalisen tiedon tarpeellisuudelle piileekin juuri tässä: havaittujen säännönmukaisuuksien tunteminen riittää ilmiöiden ennustamiseen (kunhan oletetaan, että tarkkailtava systeemi säilyy rakenteellisesti samanlaisena), mutta jos haluamme puuttua tarkkailtavan systeemin toimintaan, tarvitsemme tietoa säännönmukaisuuksien lisäksi niitä tuottavista kausaalisista suhteista. Taloutta, kansanterveyttä ja oppimistuloksia voi periaatteessa ennustaa päättelemällä puhtaan induktiivisesti havaintoihin perustuvista tilastoista tulevaisuuteen, mutta näistä tilastoista ei sellaisenaan voi päätellä, mitä tapahtuisi, mikäli talouden, terveydenhuollon tai opetusjärjestelmän rakenteita muutettaisiin. Yhteiskunnallisiin rakenteisiin kohdistuvien interventioiden seurauksista voi tehdä valistuneita arvauksia vain nojaamalla yhteiskunnallisia ja sosiaalisia mekanismeja koskeviin oletuksiin. Esimerkiksi pelkkä havainto siitä, että koulujen pieni luokkakoko korreloi parempien oppimistulosten kanssa, ei vielä kerro, voidaanko oppimistuloksia parantaa pienentämällä ryhmiä vai onko näiden muuttujien korrelaatio seurausta yhteisistä syistä.

Parhaimmillaan kokeellinen menetelmä on tehokkain ja luotettavin kausaalisen tiedon tuottamisen tapa, koska hyvä koe toimii mahdollisimman tarkasti intervention tavoin. Hyvin suunnitellussa kokeessa kokeen suorittaja vaikuttaa vain ja ainoastaan oletettuun syytekijään ja pyrkii kontrolloimaan kaikkia muita

tekijöitä, jotka saattavat vaikuttaa oletettuun seuraukseen. Näin kohteessa havaittujen muutosten voidaan päätellä olevan yksinomaan manipuloidun tekijän kausaalisia seurauksia.¹⁴ Jos kaikkia mahdollisia vaikuttavia tekijöitä ei voida kontrolloida, voidaan koetulosta vääristävät vaikutukset pyrkiä eliminoimaan satunnais-*tamalla*. Tällöin koeyksiköt jaetaan satunnaisesti koeryhmään ja verrokkiryhmään. Tarkoituksena on, että kaikki lopputulokseen mahdollisesti vaikuttavat erot yksilöiden välillä jakautuvat yhtäläisesti koeryhmään ja verrokkiryhmään eivätkä näin häiritse kausaalisen vaikutuksen arvioimista. Jos kokeentekijä itse valikoisi kohteet eri ryhmiin tai koehenkilöt voisivat itse vaikuttaa siihen, kumpaan ryhmään he kuuluvat, saattaisivat tähän valikoitumisprosessiin vaikuttavat tekijät hyvinkin korreloida sen kanssa, kuinka paljon manipulaatio kuhunkin koeyksikköön vaikuttaisi tai miten tarkastelun kohteena oleva vaikutusmuuttuja olisi käyttäytynyt ilman manipulaatiota. Tällöin puhutaan *valikoitumisvinoumasta*. Jos esimerkiksi lääketieteellisen kokeen osallistujia ei jaettaisi satunnaisesti koe- ja verrokkiryhmiin, voisi käydä niin, että koeryhmään hakeutuisi sairaampia potilaita kuin verrokkiryhmään. Tällöin havaittu koe- ja verrokkiryhmän välinen ero ei tietenkään antaisi luotettavaa arviota hoidon tehokkuudesta.

Taustamuuttujien kontrolloiminen, niin konkreettisilla manipulaatioilla suoritettussa vakioistamisessa kuin satunnaistamisessa, on virhekontrollia, joka pienentää todennäköisyyttä, että kokeessa havaitut tulokset olisivat koeasetelman artefakteja tai satuman kauppaa eivätkä kokeessa suoritettun manipulaation kausaalisia seurauksia. Vaikka koeasetelma olisi kuinka tarkkaan mietitty ja koe tunnollisesti suoritettu, on päättely kuitenkin induktiivista ja pohjimmiltaan epävarmaa. Tästä syystä kokeiden toistaminen (replikaatio) on tiedollisesti erittäin arvokasta. On kuitenkin yleisesti tunnustettua, että tieteen palkitsemisjärjestelmät eivät kannusta tutkijoita kokeiden toistamiseen. Puhuttaessa kokeiden

toistamisesta tulee erottaa kokeen toistaminen tietyn koeasetelman täsmällisestä kopioimisesta. Kun koe *toistetaan*, testataan uudelleen tiettyä hypoteesia, jota toinen tutkija tai tutkimusryhmä on jo tutkinut aiemmin järjestetyssä kokeessa. Koeasetelman *kopioimisessa* taas sovelletaan täsmälleen samaa metodologiaa täsmälleen samaan kohteeseen (esim. populaatioon). Koeasetelman kopioimisen tiedollinen arvo rajoittuu siihen, että jo olemassa olevaan havaintoaineistoon voidaan suhtautua hieman luottavaisemmin. Koeasetelman toistaminen on tärkeämpää, sillä se lisää havaintoaineistosta ilmiöön tehtävien päätelmien luotettavuutta: onnistunut toistaminen antaa tukea päättelylle, että havaintoaineistosta päätelty tulokset eivät ole riippuvaisia koeasetelman apuoletuksista tai tietyistä kohdepopulaatiosta vaan heijastavat yleisemmän ilmiön piirteitä. Kokeiden toistaminen on ensiarvoisen tärkeää myös koetuloksen soveltuvuusalan eli ulkoisen validiteetin kartoittamisessa.

KOKEIDEN ROOLI TUTKIMUKSESSA: KOETYYPIT JA NIIDEN OMINAISUUDET

Päätely koesarjassa tuotetusta havaintoaineistosta kiinnostuksen kohteena olevan ilmiön ominaisuuksiin koskee aineiston tuottavan kausaalisen prosessin luotettavuutta. Jos prosessi on luotettava ja aineistossa havaitut systemaattiset piirteet todella heijastavat tutkittavan ilmiön piirteitä eivätkä ole prosessin tuottamia artefakteja, sanotaan koeasetelmaa *sisäisesti validiksi*. Sisäisesti validi koeasetelma siis todella testaa sitä, mitä se on suunniteltu testaamaan.

Sisäinen validiteetti ei kuitenkaan vielä yksin riitä, sillä tutkijat ovat harvoin kiinnostuneita vain siitä, mitä heidän laboratorioissaan tapahtuu. Koetulosten tulisi olla sovellettavissa myös ulkomaailmaan. Sitä, kuinka hyvin koetulos on sovellettavissa

varsinaisen koeasetelman ulkopuolella, sanotaan koeasetelman *ulkoiseksi validiteetiksi*. Sisäisen ja ulkoisen validiteetin suhde on riippuvainen siitä, millaista koeasetelmaa käytetään.

Havainnollistamme sisäisen ja ulkoisen validiteetin kysymyksiä taloustieteessä tehdyillä kokeilla. Kokeellisessa taloustieteessä, ennen kaikkea niin kutsutussa käyttäytymistaloustieteessä, suosittu koeasetelmat saattavat ulkopuolisesta näyttää perin kummallisilta. Käyttäytymistaloustieteen koeasetelmat ovat suorastaan kärjistetyn riisuttuja. Tavallisesti koehenkilöt istuvat yksin tietokoneen edessä ja pelaavat anonymisti tarkasti määriteltyjä ”pelejä” toisten koehenkilöiden kanssa. Pelin rakenne on täysin läpinäkyvä ja erittäin yksinkertainen, ja se, että pelaajat ymmärtävät pelin rakenteen, tarkistetaan yleensä ennen koetta. Panoksina käytetään pieniä summia oikeaa rahaa. Saattaa näyttää siltä, että tällaisella toiminnalla on vain hyvin vähän tekemistä todellisen sosiaalisen vuorovaikutuksen kanssa – pelaajathan ”vuorovaikeuttavat” anonymisti tietokoneiden välityksellä. Koeasetelman äärimmäinen yksinkertaistaminen voidaan kuitenkin perustella kahdella toisiinsa läheisesti liittyvällä tavalla.

Ensinnäkin, jos kokeen tehtäväksi katsotaan ennen kaikkea teorian testaaminen, koeasetelman on vastattava mahdollisimman tarkoin teoriaa. Talousteorian ennusteet taas näyttävät koskevan atomististen, omaa etuaan ajavien toimijoiden käyttäytymistä. Jos pelin rahalliset panokset eivät vastaa toimijoiden preferenssejä tai jos toimijoilla on virheellisiä uskomuksia pelin rakenteesta tai muiden pelaajien uskomuksista ja haluista, koeasetelma ei vastaa sellaista tilannetta, jota talousteoria kuvaa. Näin ollen teorian ennusteiden vastainen tulos ei voisi edes periaatteessa kumota teoriaa. Tämän ”syytä teoriaa” -argumentin valossa puhtaasti teorian testaamiseen käytetyssä koeasetelmassa sisäinen validiteetti on maksimoitu ulkoisen validiteetin kustannuksella.¹⁵ Mitä tämä tarkoittaa kokeen tulosten kannalta? Koeasetelma kyllä vastaa luotet-

tavasti tilannetta, jota taloustieteellinen teoria kuvaa, mutta tuon teorian oletukset harvoin täysin toteutuvat tosielämässä. Koska teorian kuvaama tilanne on epärealistinen, tulee koeasetelmasta todellisiin vuorovaikutustilanteisiin tehtäviin yleistyksiin suhtautua varauksellisesti.

Jos ”syytä teoriaa” -argumentti on ainoa tapa puolustaa keino-tekoiselta vaikuttavia koeasetelmia, voidaan tietenkin kysyä, onko sellaisesta yhteiskuntatieteellisestä teoriasta mitään hyötyä, joka tekee ennusteita vain tällaisissa hyvin erikoislaatuissa tilanteissa. Taloustieteellisen teorian ja kokeellisen taloustieteen lähtökohdana on kuitenkin ajatus, jonka mukaan epärealistisilta vaikuttavat stilisoidut mallit ja kokeet eivät suinkaan pyri kuvaamaan konkreettisia sosiaalisia vuorovaikutustilanteita kokonaisuudessaan vaan ne pyrkivät eristämään perustavia *tendenssejä* ihmisen päätöksenteossa ja käyttäytymisessä. Ajatuksena on, että jokapäiväisessä sosiaalisessa elämässä nämä tendenssit näyttäytyvät puhtaina vain harvoin, sillä toimintamme on seurausta useista yhtä aikaa toimivista vaikuttimista. Perustavat tendenssit ovat kuitenkin koko ajan läsnä ikään kuin toiminnan rakennuspalikoina, ja yhteiskuntatieteellisen teorian on tarkoitus luoda selvyttä sosiaalisesta toiminnasta pilkkomalla se tällaisiin perustaviin osatekijöihin. Teoriaa testaava koe on tällöin klassinen laboratorio-koe, jossa kokeen tekijä pyrkii parhaansa mukaan poistamaan tai ainakin kontrolloimaan kaikkia mahdollisia häiriötekijöitä ja vaikuttamaan mahdollisimman tarkasti vain tutkimuksen kohteena olevaan muuttujaan. Tämän tulokinnan mukaan välttämätön konflikti sisäisen ja ulkoisen validiteetin välillä on siis vain näennäinen, koska tutkittu tekijä säilyttää kausaalisen tendenssinsä myös laboratorion ulkopuolella, vaikka sen vaikutukset saattavatkin peittyä muun sosiaalisen elämän hälyyn.

Taloustieteellisen teorian on usein tulkittu esittävän, että ihmisen sosiaalisen käyttäytymisen perustavimpana motivaationa on

omien intressien ajaminen. Arkikokemuksemme tuntuu kuitenkin sotivan tätä näkemystä vastaan, ja yhteiskunnan monien perusrakenteiden olemassaoloa on hankala perustella pelkäämään oman edun tavoittelulla. Myös laboratoriokokeet näyttävät osoittavan tämän oletuksen vääräksi. Esimerkiksi niin kutsutussa uhkavaatimuspelissä (*ultimatum game*) koehenkilöt (yleensä opiskelijat) pelaavat tietokoneen välityksellä peliä, jossa on kaksi roolia. Pelaajalle 1 annetaan rahasumma (joka maksetaan jälkikäteen), jonka hän voi jakaa haluamallaan tavalla pelaajan 2 kanssa. Saatuaan pelaajan 1 tarjouksen rahanjaosta pelaaja 2 voi päättää, hyväksyykö vai hylkääkö hän sen. Jälkimmäisessä tapauksessa kumpikaan pelaajista ei saa pennin hyrrää. Pelaajat eivät voi tunnistaa toisiaan pelin aikana tai sen jälkeen, ja talousteorian rationaalisen ennusteen mukaisesti pelaajan 1 pitäisi tarjota pelaajalle 2 pienin mahdollinen osuus ja pelaajan 2 tulisi hyväksyä kaikki nollasta poikkeavat tarjoukset. Sillä, että pelaajat ovat vuorovaikutuksessa anonymisti tietokoneen välityksellä ja yleensä vain kerran kutakin pelaajaa kohden, on tarkoitus varmistaa, että havaitun käyttäytymisen aiheuttavat puhtaasti pelaajien lopputulemia koskevat preferenssit. Ajatuksena siis on, että mikäli pelaaja 1 jakaa osuuden rahasta pelaajalle 2, on tämä seurausta vain ja ainoastaan siitä, että hän todella haluaa pelaajan 2 saavan jotain, eikä esimerkiksi (pohjimmiltaan itsekkäästä) häpeän tai kiusallisten sosiaalisten tilanteiden välttelmisestä tai pitkän aikavälin luottamuksen (pohjimmiltaan itsekkäästä) pohjustamisesta.

Laboratoriossa suoritettujen uhkavaatimuspelikokeiden osoittavat, että ihmiset näyttävät toimivan talousteorian ennusteiden vastaisesti. Me sekä jaamme omasta osuudestamme vapaaehtoisesti että rankaisemme muita pelaajia epäreilusta käyttäytymisestä hylkäämällä liian alhaisia tarjouksia, vaikka menettämmekin tuollain osuutemme pelaajan 1 tarjouksesta. Erään tulkinnan mukaan kokeet osoittavat, että ihmisiä motivoivat oman edun lisäksi

toisiin ihmisiin tai jonkinlaiseen reiluuteen kohdistuvat *sosiaaliset preferenssit*. Teorian mukaan sosiaaliset preferenssit ovat vakaita psykologisia tendenssejä, jotka osittain selittävät epäitsekkästä käyttäytymistä myös laboratorion ulkopuolella. Ajatuksena on, että laboratorion hyvin luonnon ja varsinaisesta sosiaalisesta vuorovaikutuksesta vapaa ympäristö mahdollistaa näiden perustavien tendenssien ilmenemisen ja suoranaisten mittaamisen. Sitä, kohdistuvatko sosiaaliset preferenssit toisen pelaajan saamiin rahamääriin (puhdas altruismi) vai yleisesti tulonjaon reiluuteen, pyritään testaamaan varioimalla pelin kannustinrakennetta, saatavissa olevaa informaatiota ja jopa käyttämällä aivokuvantamistekniikoita.

Sosiaalisten preferenssien teoria on vaikutusvaltainen mutta myös kiistanalainen tulkinta uhkavaatimuspeleissä ja muissa vastaavissa koeasetelmissa havaitusta käyttäytymisestä. Huolimatta siitä, uskommeko, että koetulokset todella kertovat jotakin valintakäyttäytymistä tuottavista psykologisista mekanismeista, kokeet ovat paljastaneet ainakin mielenkiintoisia laboratorioilmiöitä ihmisen käyttäytymisessä. Peliteoreettisten koeasetelmien ei enää yleisesti katsotakaan testaavan rationaalisen valinnan teoriaa vaan pikemminkin induktiivisesti tutkivan inhimillisen valintakäyttäytymisen säännönmukaisia piirteitä. Francesco Guala on myös ehdottanut, että hyvin pitkälle standardoitujen ja kontrolloitujen koeasetelmien voidaan katsoa olevan mittausvälineitä: toistettaessa standardoitu koe erilaisissa kulttuureissa voidaan tulosten systemaattisten erojen katsoa heijastavan jotakin eroavuutta verrattavissa kulttuureissa.¹⁶ Toisaalta, jotta voitaisiin tehdä päätelmiä siitä, millaisia erot tarkalleen ottaen ovat, on omaksuttava jokin teoria siitä, mitkä tekijät ohjaavat käyttäytymistä peliteoreettisissa kokeissa. Havaitut erot tuloksissa voivat johtua esimerkiksi erilaisista tavoista jäsentää koetilanne tai

erilaisista oikeudenmukaisuutta, kunniaa tai toisten hyvinvointia koskevista arvostuksista.

On kuitenkin avoin kysymys, miltä osin ihmisen sosiaalinen käyttäytyminen on pilkottavissa laboratoriossa tutkittaviksi tendensseiksi. Esimerkiksi kokeellisen taloustieteen kohdalla on epäselvää, voidaanko rahallisten kannustimien, anonymiteetin ja läpinäkyvän pelirakenteen katsoa poistavan tilanteen sosiaalisen hahmottamisen häiriövaikutukset, joiden ajatellaan naamioivan ”puhtaiden” sosiaalisten preferenssien ilmenemistä. Lienee selvää, että koehenkilöt joutuvat joka tapauksessa tukeutumaan johonkin tapaan hahmottaa koetilanne sosiaalisena kanssakäymisenä. Guala onkin esittänyt, että edes pitkälle idealisoituja laboratoriokokeita ei voida tulkita ilman jonkinlaista ymmärrystä niistä sosiaalisista kehyksistä, jotka ohjaavat koehenkilöiden käyttäytymistä laboratoriossa.¹⁷ Tämä on yksi tärkeä peruste sille, miksi laboratorio- ja kenttäkokeiden sekä havainnoivan ja tulkitsevan tutkimuksen tulisi parhaassa tapauksessa toimia vastavuoroisessa vuorovaikutuksessa toistensa kanssa.

Esimerkkinä tällaisesta menetelmien vuoropuhelusta voidaan tarkastella antropologi Joseph Henrichin johtamaa laajaa tutkimusta kulttuurien välisistä eroista uhkavaatimuspelissä. Henrichin tutkimusryhmä löysi huomattavia kulttuurien välisiä eroja ihmisten tavoissa jakaa palkkioita ja rangaista epäreilusta tarjouksesta. Joissakin kulttuureissa koehenkilöt jopa ristiriitaisesti hyväksyivät epäreiluja tarjouksia ja rankaisivat liiallisesta epäitsekyydestä. Heijastavatko nämä kulttuurien väliset erot laboratorio-käyttäytymisessä todella eroja perustavissa asenteissa sosiaalisesta oikeudenmukaisuudesta? Gualan mukaan näin ei välttämättä ole, sillä havaitut erot saattavat johtua kokeen osallistujien eroavista tavoista jäsentää koetilanne ennalta tuttujen sosiaalisten käytäntöjen kautta. Esimerkiksi Uuden-Guinean au- ja gnau-heimojen jäsenet hylkäsivät ylitsevuotavan anteliaat tarjoukset, koska

heidän yhteisöissään vallitsee vahva lahjojen vastavuoroisuuden normi ja koehenkilöt varoivat näin sitoutumista liian suureen ”velkasuhteeseen”.¹⁸

Vaikka sosiaalisten preferenssien teorian mukainen tulkinta kokeille on siis kiistanalainen, se edustaa hyvin klassisen laboratoriokekeen ideaa yhteiskuntatieteellisessä kontekstissa. Klassisen laboratoriokekeen ajatuksena oli kontrolloida kaikkia mahdollisia tilanteessa vaikuttavia merkittäviä tekijöitä, ja löydöksen ulkoinen validiteetti perustui oletukselle, että kontrollin avulla eristetty tendenssi on vakaa myös laboratorion ulkopuolella. Toinen mahdollisuus saavuttaa ulkoista validiteettia on suorittaa koe siinä ympäristössä, johon tietoa on tarkoitus soveltaa. *Kenttäkokeessa* koehenkilöt toimivat luonnollisessa ympäristössään ja kokeen tekijä on suorasti tai epäsuorasti vastuussa koehenkilöihin kohdistuvasta ulkoisesta manipulaatiosta.¹⁹ Koska luonnollisessa ympäristössä tutkija ei voi kontrolloida kaikkia taustamuuttujia, on satunnaistaminen paras tapa huolehtia siitä, etteivät havaitsemattomat tekijät sotke kausaalista päättelyä. Tutkija siis jakaa tutkimuskohteet satunnaisesti (mahdollisesti useampaan) manipuloitavaan ryhmään ja verrokkiryhmään. Kohteet ovat yleensä ihmisiä, mutta ne voivat myös olla suurempia yksiköitä, kuten ryhmiä, organisaatioita tai esimerkiksi kyliä.

Joissakin tapauksissa kenttäkokeet ovat varsin paikallisia ja niihin liittyvät interventiot hyvin rajattuja. Esimerkiksi vähemmistöihin kohdistuvaa syrjintää on tutkittu vastaamalla työpaikka-ilmoituksiin hakemuksilla, jotka ovat muuten täysin identtisiä mutta joissa hakijoiden nimiä on manipuloitu siten, että ne toimivat vihjeinä hakijoiden etnisistä taustoista. Yhdysvalloissa tehdyissä tutkimuksissa on kiistatta käynyt ilmi, että mikäli nimi viittaa hakijan olevan tummaihoisen, hänet kutsutaan selvästi harvemmin työhaastatteluun kuin täysin vastaavan hakemuksen lähettänyt työnhakija, jonka nimi vihjaa hakijan olevan valkoinen.²⁰

Institutionaalisten ja hallinnollisten tekijöiden vaikutusta tutkimuksissa kenttäkokeissa taas tarkastellaan kohteiden käyttäytymistä laajoissa yhteiskunnallisissa prosesseissa. Tutkimuksen kohteena voi olla vaikkapa luokkakoon vaikutus oppimistuloksiin koulutuksen varhaisvaiheissa tai kysymys siitä, kuinka valvovan viranomaisen tarkastuskäyntien määrä vaikuttaa korruption esiintymiseen tienrakennushankkeissa Indonesiassa.²¹ Usein tämän tyyppisessä tutkimuksessa tutkijat eivät itse ole vastuussa intervention toteuttamisesta vaan toimivat yhteistyössä sen toteuttavan instituution kanssa. Hallinnollisia muutoksia tutkiville kenttäkokeille on myös tyypillistä, että koeryhmän ja verrokkiryhmän erottava interventio on varsin monimutkainen eikä useinkaan ole lähtökohtaisesti selvää, millaisten mekanismien ja prosessien kautta se vaikuttaa tutkittuihin.

Toisin kuin edellä esitellyt laboratoriokokeet kenttäkokeet eivät tyypillisesti pyri eristämään vakaaksi oletettua tendenssiä. Koska ne eivät pyri eristämään tiettyä vaikutusmekanismia vaan ainoastaan arvioimaan tutkittavan tekijän *kausaalista vaikutusta* tietyssä kontekstissa, ei niiden myöskään tarvitse olla yhtä teoria-vetoisia kuin laboratoriokokeiden. Osittain tästä syystä kenttäkokeita onkin silloin tällöin arvosteltu siitä, että ne ovat liian kaukana varsinaisesta yhteiskuntatieteellisestä tutkimuksesta ja että niissä pikemminkin koeajetaan politiikkasuosituksia – vaikka hyvin suunniteltu kenttäkoe on tietenkin tiedollisesti paljon arvokkaampi kuin summamutikassa kokeilukulttuurin nimissä suoritettu ”uudistus”. Kenttäkokeen ulkoisen validiteetin kysymys koskeekin siitä, kuinka luotettavasti jokin havaittu paikallinen kausaalinen vaikutus siirtyy erilaisiin konteksteihin ja populaatioihin. Tällöin puhutaan vaikutuksen *ekstrapoloinnista*. On huomattava, että vaikutuksen ekstrapoloituvuutta on vaikea arvioida ilman ymmärrystä vaikutuksen välittävästä mekanismista, joten

täysin epäteoreettisten kenttäkokeiden yleistettävyyks on paljolti uskon varassa.²²

KOKEELLISEN IHMISTIETEEN ERITYISKYSYMYKSIÄ

Toistaiseksi emme ole juurikaan tuoneet esille seikkoja, jotka ratkaisevasti erottaisivat kokeellisen ihmis- ja yhteiskuntatieteen luonnontieteistä. Miksi sitten kokeellisen menetelmän asema on erityisesti yhteiskuntatieteissä niin paljon marginaalisempi kuin luonnontieteissä?

Useat ihmistieteiden erityislaatuista puoltavat argumentit näyttäisivät puhuvan kokeiden mielekkyyttä vastaan. Erään tällaisen ajatuskulun mukaan ihmisten toiminta on autonomista – vapaan tahdon alaista – joten käyttäytymistä koskevien yleisten lainalaisuuksien kokeellinen tutkiminen on hukkaan heitettyä aikaa. Lisäksi toisinaan ajatellaan, että ihmistieteiden päämäärä on sosiaalisten ilmiöiden merkitysten holistinen ymmärtäminen. Koska keinotekoiset koeasetelmat nimenomaan rikkovat sosiaalisen maailman (holistisen) rakenteen ja riisuvat koehenkilöiden toiminnasta kaiken mielekkyyden, ei koetuloksilla tällöin voi olla ulkoista validiteettia. Ihmistieteellisten kokeiden tulokset kuitenkin kumoavat empiirisesti ainakin ensimmäiseen argumenttiin liittyvän mielikuvan: ihmisten toiminta kokeissa on huomattavan vakaata ja jopa ennakoitavaa. Laboratoriokokeet ovat paljastaneet monia vakaita laboratorioilmiöitä (kuten konformismi, preferenssien kääntyminen, altruistinen rankaiseminen, erilaiset virittämiskokeet (*priming*), jne.), joiden tulkinnasta ja ulkoisesta validiteetista voidaan toki kiistellä mutta joiden olemassaoloa ei ole mielekästä kiistää.²³ Ajatus ihmisten autonomisuuden ja käyttäytymisen ennustamattomuuden tiiviistä yhteydestä vaikuttaakin pikemmin käsitteelliseltä kömmähdykseltä kuin yhteiskuntatieteen syvälliseltä lähtökohdalta (ks. Kokkosen luku).

Mikäli yhteiskuntatieteiden ainoana päämääränä todella pidettäisiin eläytyvän ymmärryksen avulla tuotettuja rikkaita kuvauksia ja tulkintoja, kokeellisen menetelmän käyttö ei olisi mielekästä. Sosiaalisten ilmiöiden merkitys on aina holistista, ja on vaikea nähdä, kuinka kokeelliset interventiot voisivat rikastuttaa hermeneuttista kehää. Tämän kirjan lähtökohtana on kuitenkin se, että yhteiskuntatieteellisen tutkimuksen tulisi tähdätä myös kausaaliin tietoon. Tulkitseva tutkimusote voi olla tärkeä ja jopa välttämätön osa menetelmien kokonaisuutta, joilla tähän päämäärään tähdätään, mutta lopullisten päätelmien tulisi ylittää pelkät tulkinnot tutkimuskohteiden omista tavoista jäsentää sosiaalista todellisuuttaan. Emme useinkaan halua tietää vain sitä, kuinka asiat koetaan, vaan sen, mitä todella tapahtuu. Toisaalta, kuten edellä on esitetty, tutkimuskohteiden näkökulman ymmärtäminen on välttämätöntä tulkittaessa kokeiden tuloksia. Pitkälle idealisoitujen laboratoriokokeiden ulkoista validiteettia ei voida arvioida ilman näkemystä siitä, millaisen sosiaalisen kehyksen avulla koehenkilöt koetilanteen jäsensivät. Vastaavasti kenttäkokeiden ulkoisesta validiteetista voi esittää arvioita vain perustuen jonkinlaiseen ymmärrykseen siitä sosiaalisten instituutioiden ja normien verkostosta, jossa koe tehtiin.

Kokeellisen menetelmän hyödyllisyyden tuntuu kyseenalaiseen myös se, että useat yhteiskuntatieteellisesti mielenkiintoiset kysymykset koskevat ainakin osin jo menneitä tapahtumia. Noottikriisin vaikutusta Suomen sisäpolitiikkaan ei voi suoraan kokeellisesti arvioida, koska historiaa ei voi muuttaa. Historiallisten selitysten perustana kuitenkin on oletuksia yleisemmistä sosiaalisista ja psykologisista mekanismeista (ks. Seppälän luku tässä kirjassa), joiden kokeellinen tutkiminen on usein ainakin joiltain osin mahdollista. Ongelmalliselta saattaa vaikuttaa myös se, että useat yhteiskunnallisesti merkittävät selitykselliset muutokset, kuten ikä ja sukupuoli, ovat tutkimuskohteiden piirteitä, joita

ei voida kokeellisesti muuttaa. Kuitenkin esimerkiksi juuri sukupuolen tai iän tapauksessa mielenkiinnon kohteena ei tyypillisesti ole tietyn yksilön sukupuolen tai iän vaikutus johonkin toiseen yksilöön vaan yleisemmin populaation ikä- tai sukupuolijakautuman vaikutukset joihinkin muihin muuttujiin. Nämä jakaumat ovat erilaisia eri populaatioissa, ja niitä voidaan toisinaan (ainakin periaatteessa) kokeellisesti varioida. Lisäksi sukupuolen ja iän vaikutus välittyy useimmiten näihin ominaisuuksiin kohdistuvista asenteista, ja näitä asenteita ja niiden vaikutuksia voidaan kokeellisesti tutkia muun muassa kenttäkokeilla ja pohjustuskokeilla.

Näiden yleisten huomioiden lisäksi on syytä nostaa esiin vielä muutama nimenomaan ihmisillä kokeilemiseen liittyvä erityinen metodologinen ja tutkimuseettinen ongelma:

Saako koehenkilöille valehdella? Koehenkilöiden harhaan johtamiseen ja suoranaiseen valehteluun suhtaudutaan eri tavalla eri yhteiskuntatieteissä, ja kysymykseen liittyy niin puhtaan metodologisia kuin eettisiäkin perusteita. Taloustieteessä etenkin laboratoriokokeet pitää suorittaa niin, että varmistetaan, että koehenkilöt ymmärtävät kaikki kokeen olennaiset piirteet. Kuten edellä todettiin, taloustieteilijöille kyse ei ole (ainakaan pelkästään) rehellisyydestä vaan pyrkimyksestä varmistaa, että koehenkilöiden näkemys kokeesta on se, jota taloustieteilijä haluaa tutkia tai testata, toisin sanoen, että koehenkilöiden malli vastaa tutkijan mallia. Lisäksi koehenkilöiden sumuttaminen voi tietyssä mielessä pilata koehenkilön soveltuvuuden tuleviin kokeisiin, ja tämä saattaa olla ongelma tilanteissa, joissa saatavilla oleva koehenkilöiden joukko on rajallinen. Läpinäkyvyyden vaade kuitenkin sulkee pois lukuisia tehokkaita koeasetelmia. Esimerkiksi edellä kuvattu työhaussa havaittavan syrjinnän tutkiminen kenttäkokeella tulisi mahdottomaksi. Myöskään sosiaalipsykologiset virittämiskokeet ja klassiset etnometodologiset norminrikkomiskokeet eivät tulisi kysymykseen.²⁴ Usein tällaisissa kokeissa onkin välttämätöntä

ainakin pimittää jotain relevanttia tutkimukseen liittyvää tietoa tai toisinaan suorastaan valehdella tutkimuskohteille.

Onko oikein käyttää verrokkiryhmiä tapauksissa, joissa on syytä uskoa, että toimenpide on kohteille hyödyllinen? Tämä kysymys nostetaan usein esiin lääketieteellisessä tutkimuksessa, mutta sama eettinen perusongelma koskee myös yhteiskuntatieteellisiä kokeita. Millä oikeudella vain osa koululuokista saa apuopettajan, ja millä oikeudella vain joihinkin alueen kylistä arvotaan malaria-tartuntoja ehkäisevät hyttysverkot? Kysymykseen ei ole yksiselitteistä ratkaisua, vaan verrokkiryhmän menettämää odotettua hyötyä ja toisaalta saavutetun tutkimustiedon luotettavuuden yleismaailmallista hyötyä tulee verrata tapauskohtaisesti. Tämän eettisen kysymyksen lisäksi populaation jakamiseen tutkimus- ja kontrolliryhmään liittyy puhtaasti metodologisia haasteita, koska yhteiskunnallisten ”toimenpiteiden” tapauksessa ei yleensä ole mahdollista käyttää plaseboa ja toimenpiteen kohdentaminen toiselle ryhmälle muuttaa usein myös ilman toimenpidettä jääneen kontrolliryhmän käyttäytymistä (ns. *spillover*-vaikutus).

Missä kulkevat ihmisten manipuloinnin rajat? On selvää, että esimerkiksi Milgramin tottelevaisuuskokeen kaltaisen henkisen ahdistuksen tietoinen aiheuttaminen on vähintäänkin arveluttavaa. Kuitenkin esimerkiksi korruption kokeellisessa tutkimuksessa on käytetty interventioita, jotka suoranaisesti kannustavat lahjuksien antamiseen tai vastaanottamiseen tai pienten rötösten tekemiseen.²⁵ Jälleen kerran vaakakupissa pitää samaan aikaan olla koepopulaatiolle kohdistuva haitta ja kokeesta pääteltävän kausaalisen tiedon mahdollinen hyöty. Kaikissa edellä mainituissa tapauksissa, joissa on mahdollista, että koe aiheuttaa siihen osallistuville henkilöille tai sivullisille jonkinlaista haittaa, koeasetelma tulee ennen koetta hyväksyttää tutkimusyksikön eettisellä komitealla.

Kokeellinen menetelmä on suoriin ja luotettavin tapa tuottaa tietoa kausaalisuhteista – ainakin ideaalitapauksessa. Hyvin suunniteltu ja suoritettu koe paljastaa suoraan, mitä kausaalisia seurauksia tutkimuskohteen kokeellisella manipulaatiolla on, koska se muuttaa suoraan vain ja ainoastaan tutkimuksen kohteena olevaa tekijää tavalla, joka on itsessään riippumaton tutkittavan järjestelmän muista tekijöistä tai mahdollisista ulkoisista yhteisistä syistä. Tämä peruseriaate pätee täysin samalla tavalla niin ihmisiin kuin luonnontieteisiin. Niin kontrolloitujen laboratoriokokeiden kuin kenttäkokeidenkin yleistymistä yhteiskuntatieteissä voidaankin pitää tervetulleena kehityksenä. Kausaliteetin ja täydellisen kokeen välisestä lähes käsitteellisestä yhteydestä ei kuitenkaan tule hätäisesti päätellä, että kokeellinen menetelmä olisi ainoa, tai aina edes paras, tapa tuottaa kausaalista tietoa yhteiskuntatieteissä. Useimpia yhteiskuntatieteellisesti mielenkiintoisia kausaalisia kysymyksiä ei yksinkertaisesti voi suoraan testata kokeellisesti, ja kokeiden yksioikoinen priorisointi muun yhteiskuntatieteellisen tutkimuksen kustannuksella rajoittaisi haitallisesti yhteiskuntatieteellisten tutkimuskysymysten piiriä. Kokeet ovat yksi, mutta eivät missään nimessä ainoa, tapa tuottaa kausaalista tietoa.

Perehtymistä varten

Hyviä koontiartikkeleita kokeista eri yhteiskuntatieteissä ovat esimerkiksi Banerjee & Duflo 2009 sekä Michell & Cox 2013. Baldassarri ja Abascal (2017) esittelevät erilaisia kenttäkokeita yhteiskuntatieteissä. Bardsley ym. 2009 on oivaltava, tosin voimakkaasti kantaa ottava esitys kokeellisen taloustieteen metodologiasta. Cartwright & Hardie 2012 on helppotajuinen tieteenfilosofinen esitys näyttöön perustuvasta politiikasta ja satunnaiskokeiden ongelmista. Salganik 2017 esittelee digitaalisten *online*-koeasetelmien uusia mahdollisuuksia sekä niiden eettisiä ja metodologisia ongelmia. Kirja sisältää myös katsauksen yhteiskuntatieteellisten tutkimuseettisten käytäntöjen historiaan. Petri Ylikosken kokeellista sosiologiaa käsittelevä artikkeli (2015) on erinomainen suomenkielinen lähde.

Kokeellinen yhteiskuntatiede

1. Wilson 2002; Watts 2011.
2. Sosiaalipsykologian kokeellisen tutkimuksen historiasta ks. esim. Ross & Nisbett 1991.
3. Milgram 1974.
4. Haney ym. 1973.
5. Ks. esim. Blum 2018.
6. Harrison & List 2004; Druckman & Lupia 2012; Jackson & Cox 2013.
7. Salganik 2017.
8. Salganik ym. 2006.
9. Esim. Jackson & Cox 2013.
10. Reich 2012.
11. Bogen & Woodward 1988.
12. Mayo 1996.
13. Correll ym. 2017.
14. Tämän kausaalipäätelyn periaate tunnetaan myös John Stuart Millin teoksessaan *A System of Logic* (1843, 455) esittämänä eron menetelmänä (*method of difference*): mikäli kahden samanlaisen tilanteen välillä on vain yksi ero ja ilmiö x havaitaan vain toisessa tilanteista, voidaan päätellä, että tuo ero on ilmiön x syy.
15. Bardsley ym. 2009.
16. Guala 2008.
17. Guala 2012.
18. Henrich ym. 2005, 811.
19. Harrison & List 2004; Banerjee & Duflo 2009.
20. Bertrand & Mullainathan 2004.
21. Mosteller 1995; Olken 2007.
22. Marchionni & Reijula 2018.
23. Hiljattain useat tunnetut pohjustustutkimukset on kuitenkin asetettu kyseenalaiseksi, koska niiden tulokset eivät ole olleet luotettavasti toistettavissa (ks. esim. Open Science Collaboration 2015).
24. Vrt. esim. Milgram ym. 1986.
25. Humphreys & Weinstein 2009.